



Murdoch
UNIVERSITY

MURDOCH RESEARCH REPOSITORY

This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination.

The definitive version is available at

<http://dx.doi.org/10.1016/j.csda.2004.04.014>

Lukas, M.A. and Shi, M. (2006) Sensitivity analysis of constrained linear L1 regression: Perturbations to constraints, addition and deletion of observations. Computational Statistics and Data Analysis, 51 (2). pp. 1213-1231.

<http://researchrepository.murdoch.edu.au/11456/>

Copyright: © 2006 Elsevier B.V.

It is posted here for your personal use. No further distribution is permitted.

Sensitivity Analysis of Constrained Linear L_1 Regression: Perturbations to Constraints, Addition and Deletion of Observations

Mark A. Lukas^a, Mingren Shi^b

^a Mathematics and Statistics, Murdoch University, Murdoch, W.A. 6150, Australia

Fax: 61893606332 E-mail: M.Lukas@murdoch.edu.au

^b Mathematics and Computing, University of Southern Queensland, Toowoomba, Qld. 4350, Australia

Fax: 61746315550 E-mail: shi@usq.edu.au

Abstract. This paper extends the direct sensitivity analysis of Shi and Lukas (2005) of linear L_1 (least absolute deviations) regression with linear equality and inequality constraints on the parameters. Using the same active set framework of the reduced gradient algorithm (RGA), we investigate the effect on the L_1 regression estimate of small perturbations to the constraints (constants and coefficients). It is shown that the constrained estimate is stable, but not uniformly stable, and in certain cases it is unchanged. We also consider the effect of addition and deletion of observations and determine conditions under which the estimate is unchanged. The results demonstrate the robustness of L_1 regression and provide useful diagnostic information about the influence of observations. Results characterizing the (possibly non-unique) solution set are also given. The sensitivity results are illustrated with numerical simulations on the problem of derivative estimation under a concavity constraint.

Key words: sensitivity analysis, stability, L_1 regression, least absolute deviations, robustness, diagnostics, active set, reduced gradient algorithm.

AMS (MOS) subject classifications: 62J05, 90C31, 65K05

Abbreviated title: Sensitivity of L_1 regression

1. Introduction

Consider a linear model $y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$, $i = 1, \dots, n$, where the $p \times 1$ parameter vector $\boldsymbol{\beta}$ is known to satisfy additional linear equality constraints $\mathbf{x}_i^T \boldsymbol{\beta} - y_i = 0$ and/or inequality constraints $\mathbf{x}_i^T \boldsymbol{\beta} - y_i \leq 0$. Important applications occur in parametric (and nonparametric) curve and surface fitting, and in the estimation of solutions of ill-posed and inverse problems from noisy data (see Wahba (1990)). In many such problems there is some extra information about the unknown curve or solution that can be used to constrain the parameters. In particular, if the solution (or some linear functional, e.g. an integral) is known to have a certain fixed value at some point, we obtain an equality constraint on the parameters. If the solution is known to be positive, monotone, concave or convex, this leads to certain linear inequality constraints on the parameters (see Wahba (1982) and O’Leary and Rust (1986)). Constrained regression problems also arise in certain biometric and econometric models (see Judge et al. (1985)).

Using the method of least absolute deviations, we define the constrained L_1 regression estimate of $\boldsymbol{\beta}$ to be the solution of the problem (denoted LL_1)

$$\text{minimize } S(\boldsymbol{\beta}) = \sum_{i=1}^n |\mathbf{x}_i^T \boldsymbol{\beta} - y_i|, \quad \boldsymbol{\beta} \in \mathbb{R}^p, \quad (1.1A)$$

$$\text{subject to } \mathbf{x}_i^T \boldsymbol{\beta} - y_i = 0, \quad i \in \mathcal{E} = \{n+1, \dots, n+n_{\mathcal{E}}\}, \quad (1.1B)$$

$$\mathbf{x}_i^T \boldsymbol{\beta} - y_i \leq 0, \quad i \in \mathcal{I} = \{n+n_{\mathcal{E}}+1, \dots, n+n_{\mathcal{E}}+n_{\mathcal{I}}\}, \quad (1.1C)$$

where \mathcal{E} and \mathcal{I} refer to equalities and inequalities, respectively, and we assume that $n_{\mathcal{E}} < p < n+n_{\mathcal{E}}+n_{\mathcal{I}}$. Shi and Lukas (2005) investigated the sensitivity of the solution to (1.1) with respect to perturbations in the

responses y_i and row vectors \mathbf{x}_i , $1 \leq i \leq n$, of the design matrix \mathbf{X} . The analysis was done using the active set framework of the reduced gradient algorithm (RGA) developed in Shi and Lukas (2002). In this paper we use the same framework and extend the sensitivity analysis to cover perturbations to the constraints (both the constants and coefficients) and the addition and deletion of observations.

It is well known that unconstrained L_1 regression is robust; the L_1 estimate is resistant to outliers in the response \mathbf{y} and may not be changed at all by some large perturbations in \mathbf{y} . The same is true for constrained L_1 regression, and the estimate is stable, but not uniformly stable, with respect to small perturbations in \mathbf{y} and \mathbf{x}_i , $i = 1, \dots, n$ (Shi and Lukas (2005), Ellis (1998)). Here we show that similar results apply for perturbations to the constraints in (1.1); the constrained estimate is stable for small perturbations and it may not even change with some large perturbations in the constraints. The stability is not uniform but depends on the degree of ill-conditioning and, for the constraint coefficient vectors, also on how close the estimate is to being non-unique.

For unconstrained L_1 regression, the effect of deletion of observations has been investigated by Narula and Wellington (1985), Dupačová (1992) and Castillo et al. (2001). These works are based on the LP formulation of the L_1 regression problem. Here we use the direct active set approach on the general LL_1 problem (1.1) to find conditions under which an existing solution remains optimal after the deletion of an observation. We also show how the RGA can be used to find a new solution to the problem with a deleted observation, starting from the original solution. The L_1 version of the Cook distance can then be computed efficiently to determine the influence of each observation on the L_1 regression estimate. This provides important diagnostic information about the model. A result about the addition of an observation shows that constrained L_1 regression is robust with respect to the new response.

A brief description of the RGA framework and optimality conditions for (1.1) is given in Section 2. It is known that the solution of an L_1 regression problem can be non-unique (Bloomfield and Steiger (1983)). In Section 3 we derive results characterizing the solution set of (1.1) and show that it can be computed efficiently using the RGA. In Sections 4 and 5, we investigate the effect on the solution to (1.1) of a perturbation to the constants and coefficients of the constraints, respectively. Results relating to the deletion and addition of an observation are derived in Sections 6 and 7, respectively.

In Section 8 we consider the problem of estimating the derivative $f(t) = g'(t)$ from data $y_i = g(t_i) + \varepsilon_i$, $i = 1, \dots, n$, under the concavity constraint $f''(t) \leq 0$. The estimation of derivatives arises in many applications, in particular in the analysis of growth curves (Gasser et al. (1984) and Eubank (1988)) and pharmacokinetic response curves (Song et al. (1995)). Numerical differentiation is an ill-posed problem, meaning that the solution is sensitive to errors in the data (see Anderssen and Bloomfield (1974)), and such problems were a major motivation for this work. We use a truncated trigonometric series for the estimate $f_p(t)$ of the derivative and find the estimate by solving a constrained linear L_1 regression problem. Results of numerical simulations illustrate the sensitivity results from Sections 4, 5 and 6, and also a result from Shi and Lukas (2005) on perturbations to the responses. We also consider the effect of increasing the number of points at which the the constraint $f_p''(t) \leq 0$ is evaluated.

2. Active set framework

We use the same notation, definitions and optimality results as in Shi and Lukas (2005), but we repeat them here for completeness (see also Bloomfield and Steiger (1983), and Osborne (1985)). An inequality constraint $\mathbf{x}_i^T \boldsymbol{\beta} - y_i \leq 0$ is said to be active at $\boldsymbol{\beta}$ if $\mathbf{x}_i^T \boldsymbol{\beta} - y_i = 0$ (equality constraints are automatically

active). Similarly we say that a model equation $y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$ is an active equation at $\boldsymbol{\beta}$ if the residual is 0. Let the *active set* at a given point $\boldsymbol{\beta}$ and its associated sets be

$$\mathcal{A} = \mathcal{A}(\boldsymbol{\beta}) = \{i \mid r_i(\boldsymbol{\beta}) \equiv \mathbf{x}_i^T \boldsymbol{\beta} - y_i = 0, 1 \leq i \leq n + n_{\mathcal{E}} + n_{\mathcal{I}}\}, \quad (2.1A)$$

$$\mathcal{A}_S = \mathcal{A} \cap \{1, \dots, n\}, \quad \mathcal{A}_I = \mathcal{A} \cap \mathcal{I}, \quad \mathcal{A}_S^c = \{1, \dots, n\} \setminus \mathcal{A}_S. \quad (2.1B)$$

Note that the residual $r_i(\boldsymbol{\beta})$ is the opposite of the usual definition.

Definition 2.1 (1) Denoting $\text{rank}\{\mathbf{x}_i\}_{i=1}^{n+n_{\mathcal{E}}+n_{\mathcal{I}}} = r \leq p$, we say $\boldsymbol{\beta}$ is a *base point* of problem (1.1) if the rank of the active gradient vectors $\{\mathbf{x}_i \mid i \in \mathcal{A}\}$ is r . (2) A feasible base point $\boldsymbol{\beta}$ is *nondegenerate* if the active gradient vectors $\{\mathbf{x}_i \mid i \in \mathcal{A}\}$ are linearly independent. If every feasible base point of problem (1.1) is nondegenerate, the problem (1.1) is said to be *nondegenerate* and otherwise it is *degenerate*.

We will be using the following known optimality and uniqueness conditions.

Theorem 2.1 Necessary and sufficient conditions for $\boldsymbol{\beta}$ to be a solution to (1.1) are that $\boldsymbol{\beta}$ is feasible, i.e.

$$\mathbf{x}_i^T \boldsymbol{\beta} - y_i = 0, \quad i \in \mathcal{E}, \quad \text{and} \quad \mathbf{x}_i^T \boldsymbol{\beta} - y_i \leq 0, \quad i \in \mathcal{I}, \quad (2.2)$$

and there exist *multipliers* λ_i , $i \in \mathcal{A}$, such that

$$\mathbf{c} \equiv \sum_{i \in \mathcal{A}_S^c} \sigma_i \mathbf{x}_i = \sum_{i \in \mathcal{A}_S} \lambda_i \mathbf{x}_i + \sum_{i \in \mathcal{E}} \lambda_i \mathbf{x}_i + \sum_{i \in \mathcal{A}_I} \lambda_i \mathbf{x}_i, \quad (2.3A)$$

$$|\lambda_i| \leq 1, \quad i \in \mathcal{A}_S, \quad \text{and} \quad \lambda_i \leq 0, \quad i \in \mathcal{A}_I, \quad (2.3B)$$

where $\sigma_i = \sigma_i(\boldsymbol{\beta}) = \text{sign}(r_i(\boldsymbol{\beta}))$. Furthermore, (1.1) has a unique solution if and only if the inequalities (2.3B) are strict.

To illustrate this result, consider the simplest unconstrained, univariate case with $X = [1, \dots, 1]^T$. Here we have $r_i = \beta - y_i$ so c is just the number of values y_i that are less than β minus the number that are greater than β . For optimality of β , condition (2.3) requires this number c to be less than or equal to one in magnitude, i.e. $-1, 0$ or 1 . Clearly, if $c = 0$, the unique solution is the median of the data, while if $c = \pm 1$, all β between the middle two data values are optimal.

We will also use the following well known, fundamental result.

Theorem 2.2 The minimum value of problem (1.1) is achieved and can be found at a (feasible) base point of (1.1). (It is assumed that the feasible region is not empty.)

The sensitivity results will be derived in the active set framework of the reduced gradient algorithm (RGA) as developed in Shi and Lukas (2002) (see also Bloomfield and Steiger (1983) and Osborne (1985, 1987)). The RGA is a descent method which moves from one feasible base point to an adjacent one. We only consider the nondegenerate case here (see Shi and Lukas (2002) for the degenerate case). The main steps of the RGA are as follows.

Step (0) Find a feasible starting point $\boldsymbol{\beta}^{(0)}$. For an unconstrained problem, an arbitrary point $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$ (e.g. $\hat{\boldsymbol{\beta}} = \mathbf{0}$) can serve as the starting point. Then we add *artificial constraints* $\mathbf{e}_i^T \boldsymbol{\beta} - \hat{\beta}_i \equiv \beta_i - \hat{\beta}_i \leq 0$, $i \in \{1, 2, \dots, p\}$ to (1.1A) to form an augmented problem for which $\hat{\boldsymbol{\beta}}$ is a base point (with rank p). For the general constrained LL_1 problem, we solve an auxiliary LP problem (see Shi and Lukas (2002)) to find a feasible starting point. At the start an LL_1 problem may also be augmented with artificial constraints. The active set is $\hat{\mathcal{A}} = \hat{\mathcal{A}}_S \cup \mathcal{E} \cup \hat{\mathcal{A}}_I \cup \hat{\mathcal{A}}_0$, where $\hat{\mathcal{A}}_S$ and $\hat{\mathcal{A}}_I$ are defined in (2.1) and $\hat{\mathcal{A}}_0$ denotes a set of indices of artificial constraints. Note that these artificial constraints are only a computational device; it will be seen in step (3) that they are not enforced by the algorithm. Set $k = 0$ and go to step (1).

Step (1) Compute the “cost” vector and *multipliers*. Let $\boldsymbol{\beta}^{(k)} = \widehat{\boldsymbol{\beta}}$ and let its active set, the *base matrix* \widehat{A} and its *inverse* \widehat{D} be, respectively,

$$\widehat{A} = \{j_1, j_2, \dots, j_p\}, \quad \widehat{A} = [\mathbf{x}_{j_1} \ \mathbf{x}_{j_2} \ \dots \ \mathbf{x}_{j_p}], \quad \widehat{D} = \widehat{A}^{-1} = [\widehat{\mathbf{d}}_1 \ \widehat{\mathbf{d}}_2 \ \dots \ \widehat{\mathbf{d}}_p]^T,$$

where $\widehat{\mathbf{d}}_i^T$ is the i th row vector of \widehat{D} . As in (2.3A), compute the “cost” vector $\widehat{\mathbf{c}}$ and the multipliers $\widehat{\lambda}_i$ by

$$\widehat{\mathbf{c}} = \mathbf{c}(\widehat{\boldsymbol{\beta}}) \equiv \sum_{i \in \widehat{\mathcal{A}}_S^c} \sigma_i \mathbf{x}_i = \sum_{i=1}^p \widehat{\lambda}_i \mathbf{x}_{j_i} \quad \text{and} \quad \widehat{\lambda}_i = \widehat{\mathbf{d}}_i^T \widehat{\mathbf{c}}, \quad i = 1, 2, \dots, p, \quad (2.4)$$

where $\sigma_i = \text{sign}(r_i(\widehat{\boldsymbol{\beta}}))$.

Step (2) Test the optimality conditions. Let

$$\max\{|\widehat{\lambda}_i| : j_i \in \widehat{\mathcal{A}}_0\} = |\widehat{\lambda}_{q_0}|, \quad \max\{|\widehat{\lambda}_i| : j_i \in \widehat{\mathcal{A}}_S\} = |\widehat{\lambda}_{q_1}| \quad \text{and} \quad \max\{\widehat{\lambda}_i : j_i \in \widehat{\mathcal{A}}_I\} = \widehat{\lambda}_{q_2}.$$

If

$$|\widehat{\lambda}_{q_0}| = 0, \quad |\widehat{\lambda}_{q_1}| \leq 1 \quad \text{and} \quad \widehat{\lambda}_{q_2} \leq 0, \quad (2.5)$$

then, from (2.3), $\widehat{\boldsymbol{\beta}}$ is optimal and the algorithm is terminated. Otherwise, go to step (3)

Step (3) Choose a feasible descent direction \mathbf{d} . Any one of the choices

$$\mathbf{d} = \begin{cases} -\text{sign}(\widehat{\lambda}_{q_0}) \widehat{\mathbf{d}}_{q_0}, & \text{if } |\widehat{\lambda}_{q_0}| \neq 0, \\ -\text{sign}(\widehat{\lambda}_{q_1}) \widehat{\mathbf{d}}_{q_1}, & \text{if } |\widehat{\lambda}_{q_1}| > 1, \\ -\widehat{\mathbf{d}}_{q_2}, & \text{if } \widehat{\lambda}_{q_2} > 0, \end{cases} \quad (2.6)$$

is a feasible descent direction at $\widehat{\boldsymbol{\beta}}$ for (1.1) (see Shi and Lukas (2002)). Note that any direction is regarded as feasible for the artificial constraints. A move from $\widehat{\boldsymbol{\beta}}$ in the direction \mathbf{d} , with corresponding index j_q , $q \in \{q_0, q_1, q_2\}$, will leave the boundary plane P_{j_q} (defined by $P_j = \{\boldsymbol{\beta} \mid \mathbf{x}_j^T \boldsymbol{\beta} - y_j = 0\}$) but remain on the intersection of the other $p - 1$ boundary planes P_{j_i} , $j_i \in \widehat{\mathcal{A}} \setminus \{j_q\}$. Hence the index j_q will become inactive. Note that more than one of the three choices of \mathbf{d} in (2.6) may be possible. But if $|\widehat{\lambda}_{q_0}| \neq 0$, we choose $q = q_0$ since we want to delete the artificial constraints.

Step (4) Compute the search step length. Let the shortest step to the boundary planes P_j associated with *inactive equations* and with *inactive inequality constraints* be, respectively,

$$\delta' = \delta'(\mathbf{d}) = \min\{\delta_j \equiv -r_j(\widehat{\boldsymbol{\beta}})/\mathbf{x}_j^T \mathbf{d} \mid \delta_j > 0, \ n \geq j \in \widehat{\mathcal{A}}_S^c, \ \mathbf{x}_j^T \mathbf{d} \neq 0\}, \quad (2.7A)$$

$$\alpha' = \alpha'(\mathbf{d}) = \min\{\alpha_j \equiv -r_j(\widehat{\boldsymbol{\beta}})/\mathbf{x}_j^T \mathbf{d} \mid \alpha_j > 0, \ n < j \in \mathcal{I} \setminus \widehat{\mathcal{A}}_I, \ \mathbf{x}_j^T \mathbf{d} > 0\}, \quad (2.7B)$$

where r_j is defined in (2.1). If $\alpha' < \delta'$, then we choose the step length $\alpha = \alpha'$. Otherwise, we choose $\alpha = \delta'$.

Step (5) Update the base point and its inverse matrix. Let $\overline{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} + \alpha \mathbf{d}$ be the new base point with base matrix \overline{A} , and let m be the index of the new active equation or constraint. The new inverse matrix $\overline{D} = \overline{A}^{-1} = [\overline{\mathbf{d}}_1 \ \overline{\mathbf{d}}_2 \ \dots \ \overline{\mathbf{d}}_p]^T$ can be obtained from \widehat{D} using the formula

$$\overline{\mathbf{d}}_i = \begin{cases} \frac{1}{\mathbf{x}_m^T \mathbf{d}} \mathbf{d} = \frac{1}{\mathbf{a}_m^T \widehat{\mathbf{d}}_q} \widehat{\mathbf{d}}_q, & i = q, \\ \widehat{\mathbf{d}}_i - (\mathbf{x}_m^T \widehat{\mathbf{d}}_i) \frac{1}{\mathbf{x}_m^T \mathbf{d}} \mathbf{d} = \widehat{\mathbf{d}}_i - (\mathbf{x}_m^T \widehat{\mathbf{d}}_i) \overline{\mathbf{d}}_q, & i \neq q. \end{cases} \quad (2.8)$$

Since $\alpha \leq \delta'$, the new value of S is

$$S(\bar{\boldsymbol{\beta}}) = S(\hat{\boldsymbol{\beta}}) + \alpha(\bar{\mathbf{c}}^T \mathbf{d} + \sum_{i \in \hat{\mathcal{A}}_S} |\mathbf{x}_i^T \mathbf{d}|). \quad (2.9)$$

Define $\boldsymbol{\beta}^{(k+1)} = \bar{\boldsymbol{\beta}}$. Set $k = k + 1$ and go to step (1).

Since there are finitely many base points and the RGA is a descent method, it terminates at an optimal base point.

3. Structure of non-unique solutions

For the unconstrained L_1 regression problem, Bloomfield and Steiger (1983) (Sec. 1.3) derived several important properties of the (possibly non-unique) solution set. For the general problem (1.1), we will characterize the solution set \mathcal{M} in the framework of the RGA and show how it can be computed easily.

It is easy to show that the function $S(\boldsymbol{\beta})$ is convex. Because the feasible region defined by the constraints (1.1B) and (1.1C) is convex, we immediately have the following result.

Lemma 3.1 The solution set \mathcal{M} of (1.1) is convex, so if \mathcal{M} contains more than one point, it contains infinitely many.

Since the solution set \mathcal{M} is convex, the same argument used in Bloomfield and Steiger (1983) (Sec. 1.3 Thm. 4) applies to give the following result (the case of inactive constraints is obvious).

Theorem 3.1 Any two points $\tilde{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\beta}}$ of \mathcal{M} are on the same side of every boundary plane corresponding to an inactive equation or constraint, i.e. $r_i(\tilde{\boldsymbol{\beta}})r_i(\bar{\boldsymbol{\beta}}) > 0$ for all $i \in \mathcal{A}^c(\tilde{\boldsymbol{\beta}}) \cap \mathcal{A}^c(\bar{\boldsymbol{\beta}})$.

To characterize the solution set, there are two cases to consider. Let $\bar{\mathbf{X}} = [\mathbf{x}_1 \cdots \mathbf{x}_{n+n_{\mathcal{E}}+n_{\mathcal{I}}}]^T$ be the extended design matrix consisting of all the row vectors for the model equations and constraints. First we consider the case where $\bar{\mathbf{X}}$ has full rank p . In this case, for any base point $\boldsymbol{\beta}$, the active gradients \mathbf{x}_i , $i \in \mathcal{A}_S(\boldsymbol{\beta}) \cup \mathcal{E} \cup \mathcal{A}_I(\boldsymbol{\beta})$, span \mathbb{R}^p . From Theorem 2.1, we know that the solution set \mathcal{M} of (1.1) contains a base point. Using the same argument as in Bloomfield and Steiger (1983) (Sec. 1.3 Thm. 5) with Theorem 3.1, we have the following.

Theorem 3.2 If $\bar{\mathbf{X}}$ has full rank, the solution set \mathcal{M} of (1.1) is the convex hull of the base points contained in it.

Thus, to find all the non-unique solutions of (1.1) it is necessary only to compute all the optimal base points. This can be done efficiently using the RGA above. First we obtain one optimal base point $\tilde{\boldsymbol{\beta}}$ and compute the corresponding multipliers $\tilde{\lambda}_i$, $i = 1, \dots, p$, $j_i \in \tilde{\mathcal{A}}$. Since $\tilde{\boldsymbol{\beta}}$ is not unique, Theorem 2.1 implies there is a multiplier satisfying either $|\tilde{\lambda}_{q_1}| = 1$ for some $j_{q_1} \in \tilde{\mathcal{A}}_S$ or $\tilde{\lambda}_{q_2} = 0$ for some $j_{q_2} \in \tilde{\mathcal{A}}_I$. Then the direction $\mathbf{d} = -\text{sign}(\tilde{\lambda}_{q_1})\tilde{\mathbf{d}}_{q_1}$ or $\mathbf{d} = -\tilde{\mathbf{d}}_{q_2}$ is a feasible direction for which the values of $S(\tilde{\boldsymbol{\beta}} + \alpha\mathbf{d})$ are constant for all sufficiently small α . With this change to step (3) of the RGA above, for each “constant” search direction we continue the algorithm to find a new optimal base point. Then, for each new point, we go to step (1) and continue, making sure that a new “constant” search direction \mathbf{d} is not simply the opposite of the previous search direction. In this way the algorithm will find all the optimal base points.

Example 3.1 minimize $S(\boldsymbol{\beta}) = |2\beta_1 + \beta_2 - 4| + |-\beta_1 + \beta_2 - 1| + |2\beta_1 - 5| + |\beta_1 + \beta_2 - 2|$.

It is easy to check that $\tilde{\boldsymbol{\beta}} = (2, 0)^T$ is an optimal base point with $\tilde{\mathcal{A}} = \{1, 4\}$ and multipliers $(\tilde{\lambda}_1, \tilde{\lambda}_2) = (0, -1)$. Since $|\tilde{\lambda}_2| = 1$, the solution is not unique and the direction $\mathbf{d} = -\text{sign}(\tilde{\lambda}_2)\tilde{\mathbf{d}}_2 = +(-1, 2)^T$ is a “constant” search direction. Moving from $(2, 0)^T$ along this direction yields the new optimal base point $(1, 2)^T$ with active set $\{1, 2\}$. The solution set is the closed line segment between $(2, 0)^T$ and $(1, 2)^T$.

In some situations it may be desirable to single out just one point from an infinite solution set. A plausible criterion is to choose (if possible) the solution that has the best balance of its residuals in that the sum is closest to zero. In Example 3.1 the residuals at the two optimal base points are $[r_i((2, 0)^T)] = [0, -3, -1, 0]$ and $[r_i((1, 2)^T)] = [0, 0, -3, 1]$, and for any optimal solution the residuals are $[r_i((2, 0)^T + \alpha(-1, 2)^T)] = [0, -3 + 3\alpha, -1 - 2\alpha, \alpha]$, $0 \leq \alpha \leq 1$. So $(1, 2)^T$ is the preferred optimal solution.

The second case to consider is where \overline{X} does not have full rank. If $\text{rank}(\overline{X}) = r < p$, we can choose r linearly independent columns and store them as \overline{X}_1 , and store the remaining columns as \overline{X}_2 . Without loss of generality, reorder the columns of \overline{X} so that $\overline{X} = [\overline{X}_1 \ \overline{X}_2]$. The matrix \overline{X}_2 is related to \overline{X}_1 by $\overline{X}_2 = \overline{X}_1 B$ for some $r \times (p - r)$ matrix B . Hence, if $\beta^T = [\beta_1^T, \beta_2^T]$, with $\beta_1 \in \mathbb{R}^r$, then

$$\overline{X}\beta = \overline{X}_1\beta_1 + \overline{X}_2\beta_2 = \overline{X}_1(\beta_1 + B\beta_2). \quad (3.1)$$

Suppose $\hat{\beta}$ is an optimal base point of (1.1) with active set $\hat{\mathcal{A}}_S \cup \mathcal{E} \cup \hat{\mathcal{A}}_I$. Without loss of generality we can assume that $\hat{\beta}_i = 0$, $i = r + 1, \dots, p$, (since from (3.1), $[(\hat{\beta}_1 + B\hat{\beta}_2)^T, \mathbf{0}^T]^T$ also solves (1.1)). Then we can consider $\hat{\beta}$ as an optimal base point of (1.1) with active artificial constraints $\beta_i \leq 0$, $i = r + 1, \dots, p$, i.e. $\hat{\mathcal{A}}_0 = \{r + 1, \dots, p\}$. With the active set $\hat{\mathcal{A}} = \hat{\mathcal{A}}_S \cup \mathcal{E} \cup \hat{\mathcal{A}}_I \cup \hat{\mathcal{A}}_0$, the base matrix and its inverse are of the form

$$\hat{A} = \begin{bmatrix} \hat{A}_1 & 0 \\ \hat{A}_2 & I \end{bmatrix} \quad \text{and} \quad \hat{D} = \begin{bmatrix} \hat{A}_1^{-1} & 0 \\ -\hat{A}_2\hat{A}_1^{-1} & I \end{bmatrix}, \quad (3.2)$$

where A_1 is $r \times r$ and nonsingular, and $I = I_{(p-r) \times (p-r)}$. Consider the problem (1.1)₁ obtained by setting the variables $\beta_i = 0$, $i = r + 1, \dots, p$, in (1.1), so that the extended design matrix is \overline{X}_1 . Since \overline{X}_1 has full rank, the solution set of (1.1)₁ is the convex hull of its optimal base points as in Theorem 3.2.

The following result shows that the solution set \mathcal{M} of (1.1) is unbounded in this case.

Theorem 3.3 Suppose $\text{rank}(\overline{X}) = \text{rank}(\overline{X}_1) = r < p$ and let $\hat{\theta}$ be any optimal base point of (1.1)₁. Then $\hat{\beta} \equiv [\hat{\theta}^T, \mathbf{0}^T]^T$ is an optimal base point of (1.1) with (after reordering) base matrix and inverse in (3.2). Moreover, for any $\kappa_j \in \mathbb{R}$, the point

$$\tilde{\beta} = [\hat{\theta}^T, \mathbf{0}^T]^T + \sum_{j \in \hat{\mathcal{A}}_0} \kappa_j \hat{\mathbf{d}}_j, \quad (3.3)$$

where $\hat{\mathbf{d}}_j^T$, $j = r + 1, \dots, p$, is the j th row of \hat{D} in (3.2), is also a solution to (1.1). If (1.1)₁ has a unique solution $\hat{\theta}$, then the solution set \mathcal{M} of (1.1) is the $(p - r)$ -dimensional affine set defined by the right hand side of (3.3) with arbitrary κ_j .

Proof As above, the matrix \overline{X}_2 is related to \overline{X}_1 by $\overline{X}_2 = \overline{X}_1 B$ for some $r \times (p - r)$ matrix B . Clearly $\hat{\beta}$ is a feasible base point for (1.1), and it is optimal since, from (3.1), $S(\hat{\beta}) = S_1(\hat{\theta}) \leq S_1(\beta_1 + B\beta_2) = S(\beta)$ for any $\beta \in \mathbb{R}^p$, where S_1 is the objective function for (1.1)₁. Restricting attention to the first r rows (corresponding to the active set) in $\overline{X}_2 = \overline{X}_1 B$, clearly we have $\hat{A}_2 = B^T \hat{A}_1$ so $B^T = \hat{A}_2 \hat{A}_1^{-1}$. Then, from \hat{D} in (3.2), $[-B^T \ I] = [\hat{\mathbf{d}}_{r+1}, \dots, \hat{\mathbf{d}}_p]^T$ and (3.3) can be rewritten as $\tilde{\beta} = [\hat{\theta}^T, \mathbf{0}^T]^T + [-B^T \ I]^T \beta_2$, where $\beta_2 = (\kappa_{r+1}, \dots, \kappa_p)^T$. Since $\overline{X}\tilde{\beta} = \overline{X}\hat{\beta}$ (from $\overline{X}_2 = \overline{X}_1 B$), it follows that $\tilde{\beta}$ is a solution to (1.1) for arbitrary $\beta_2 \in \mathbb{R}^{p-r}$.

Now suppose $\hat{\theta}$ is the unique solution to (1.1)₁ so $\hat{\beta} = [\hat{\theta}^T, \mathbf{0}^T]^T$ is an optimal base point of (1.1). If $\beta = [\beta_1^T, \beta_2^T]^T$ is any optimal solution to (1.1), then $\beta_1 + B\beta_2$ is feasible for (1.1)₁ since $\overline{X}_1(\beta_1 + B\beta_2) = \overline{X}\beta$. Furthermore it is optimal for (1.1)₁ since $S_1(\beta_1 + B\beta_2) = S(\beta) = S(\hat{\beta}) = S_1(\hat{\theta})$. Because $\hat{\theta}$ is unique, we must have $\beta_1 + B\beta_2 = \hat{\theta}$ and so $\beta = [\hat{\theta}^T, \mathbf{0}^T]^T + [-B^T \ I]^T \beta_2$, which is of the form in (3.3). ■

An example of an unconstrained problem illustrating Theorem 3.3 is discussed in Shi and Lukas (2002) (Example 5.1 and Remark 5.1). Here we give a simple constrained example.

Example 3.2 Consider the LL_1 problem

$$\begin{aligned} \text{minimize } S(\boldsymbol{\beta}) &= |\beta_1 + \beta_3| + |(\beta_1 + \beta_2 + \beta_3 + \beta_4)/2 - 1| + |(\beta_1 - \beta_2 + \beta_3 - \beta_4)/2 - 1| \\ \text{subject to } &\beta_2 + \beta_4 \leq 0 \quad \text{and} \quad \beta_1 + \beta_3 \leq 1. \end{aligned}$$

By writing down \overline{X} and letting \overline{X}_1 be the first two columns, it is obvious that $\text{rank}(\overline{X}) = \text{rank}(\overline{X}_1) = 2$ here. It is not hard to confirm that $\hat{\boldsymbol{\beta}} = \mathbf{0}$ is an optimal base point with the optimal value $S(\hat{\boldsymbol{\beta}}) = 2$ and two artificial constraints $\beta_i \leq 0$, $i = 3, 4$. Its base matrix is $\hat{A} = [\mathbf{e}_1 + \mathbf{e}_3, \mathbf{e}_2 + \mathbf{e}_4, \mathbf{e}_3, \mathbf{e}_4]$ (where \mathbf{e}_i is the i th coordinate vector), the inverse base matrix is

$$\hat{D} \equiv [\hat{\mathbf{d}}_S, \hat{\mathbf{d}}_I, \hat{\mathbf{d}}_{01}, \hat{\mathbf{d}}_{02}]^T = [\mathbf{e}_1, \mathbf{e}_2, -\mathbf{e}_1 + \mathbf{e}_3, -\mathbf{e}_2 + \mathbf{e}_4]^T,$$

and the corresponding multipliers are $(-1, 0, 0, 0)^T$. Hence $\mathbf{d} = +\hat{\mathbf{d}}_S$ and $\mathbf{d} = -\hat{\mathbf{d}}_I$ are “constant” search directions. First, considering the direction $-\hat{\mathbf{d}}_I$, we can show that the points $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \kappa_I(-\hat{\mathbf{d}}_I) + \kappa_{01}\hat{\mathbf{d}}_{01} + \kappa_{02}\hat{\mathbf{d}}_{02}$, with $0 \leq \kappa_I \leq 2$ and arbitrary κ_{01} and κ_{02} , are also solutions; $(\mathbf{e}_2 + \mathbf{e}_4)^T \tilde{\boldsymbol{\beta}} = -\kappa_I \leq 0$ and $(\mathbf{e}_1 + \mathbf{e}_3)^T \tilde{\boldsymbol{\beta}} = 0 \leq 1$ so $\tilde{\boldsymbol{\beta}}$ is feasible and $S(\tilde{\boldsymbol{\beta}}) = 0 + |-\kappa_I/2 - 1| + |\kappa_I/2 - 1| = 2 = S(\hat{\boldsymbol{\beta}})$ since $0 \leq \kappa_I \leq 2$. Note, however, if $\kappa_I > 2$, the point $\tilde{\boldsymbol{\beta}}$ is not optimal. With the same reasoning, if $0 \leq \kappa_I \leq 2$, then $(0, -\kappa_I)^T$ is a solution to the corresponding problem defined by setting $\beta_3 = \beta_4 = 0$. In particular, with $\kappa_I = 0$ or $\kappa_I = 2$, respectively, $\hat{\boldsymbol{\theta}} \equiv (0, 0)^T$ or $\hat{\boldsymbol{\theta}} \equiv (0, -2)^T$ is an optimal base point and $\tilde{\boldsymbol{\beta}}$ is of the form in (3.3).

Similarly, by considering the “constant” direction $\hat{\mathbf{d}}_S$, we can show that the points $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \kappa_S(\hat{\mathbf{d}}_S) + \kappa_{01}\hat{\mathbf{d}}_{01} + \kappa_{02}\hat{\mathbf{d}}_{02}$, with $0 \leq \kappa_S \leq 1$ and arbitrary κ_{01} and κ_{02} , are also solutions to the given problem, and $(\kappa_S, 0)^T$ solves the problem with $\beta_3 = \beta_4 = 0$. So $\tilde{\boldsymbol{\beta}}$ is also of the form in (3.3). By Lemma 3.1, any convex combination of $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ is also a solution.

4. Perturbations to the constants of constraints

In this and the next section, we will analyse a perturbation to the constraints of the LL_1 problem. Let $\tilde{\boldsymbol{\beta}}$ be an optimal base point for (1.1) (not necessarily unique). Suppose that the l th constraint, $\mathbf{x}_l^T \boldsymbol{\beta} - y_l = 0$ (or ≤ 0), is perturbed to $\mathbf{x}_l^T \boldsymbol{\beta} - (y_l + \Delta y_l) = 0$ (or ≤ 0), where $\Delta y_l \neq 0$, and denote the new problem by (1.1)^[l]. The effect of the perturbation depends on the type of constraint.

4.1. Equality constraints

In this case the point $\tilde{\boldsymbol{\beta}}$ is not feasible for the new problem (1.1)^[l]. However, a certain move from $\tilde{\boldsymbol{\beta}}$ yields a new feasible and optimal point, as shown below.

Theorem 4.1 If $l = j_q \in \mathcal{E} \subset \tilde{\mathcal{A}}$, let the new point be $\overline{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} + |\Delta y_l| \mathbf{d}$, where $\mathbf{d} = \text{sign}(\Delta y_l) \tilde{\mathbf{d}}_q$. If the perturbation satisfies $|\Delta y_l| < \min\{\delta'(\mathbf{d}), \alpha'(\mathbf{d})\}$ (see (2.7)), then $\overline{\boldsymbol{\beta}}$ is an optimal base point for the new problem (1.1)^[l]. The new optimal point and optimal value are, respectively,

$$\overline{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} + |\Delta y_l| \mathbf{d} = \tilde{\boldsymbol{\beta}} + \Delta y_l \tilde{\mathbf{d}}_q \quad \text{and} \quad S^{[l]}(\overline{\boldsymbol{\beta}}) = S(\tilde{\boldsymbol{\beta}}) + \Delta y_l \tilde{\lambda}_q. \quad (4.1)$$

Proof By direct substitution, $\overline{\boldsymbol{\beta}}$ satisfies the new equality constraint since $\mathbf{x}_l^T \mathbf{d}_q = 1$. This, together with the condition $|\Delta y_l| < \min\{\delta'(\mathbf{d}), \alpha'(\mathbf{d})\}$ and the fact that the direction \mathbf{d} satisfies $\mathbf{x}_i^T \mathbf{d} = 0$, $\forall i \in \tilde{\mathcal{A}} \setminus \{l\}$, ensures that $\overline{\boldsymbol{\beta}}$ is a feasible base point. Moreover, the active set $\overline{\mathcal{A}}^{[l]}$ for (1.1)^[l] at $\overline{\boldsymbol{\beta}}$ is the same as $\tilde{\mathcal{A}}$ and hence the inverse base matrix is not changed. Since the move does not encounter any boundary planes of inactive equations, the “cost” vector and multipliers at $\overline{\boldsymbol{\beta}}$ for (1.1)^[l] remain the same as those at $\tilde{\boldsymbol{\beta}}$ for (1.1). Thus $\overline{\boldsymbol{\beta}}$ is an optimal point for (1.1)^[l]. Since $|\Delta y_l| < \delta'(\mathbf{d})$ and $\mathbf{x}_i^T \mathbf{d} = 0$, $\forall i \in \tilde{\mathcal{A}}_S$, from (2.9) and (2.4) we obtain the new optimal value

$$S^{[l]}(\overline{\boldsymbol{\beta}}) = S(\tilde{\boldsymbol{\beta}}) + |\Delta y_l| (\tilde{\mathbf{c}}^T \mathbf{d} + \sum_{i \in \tilde{\mathcal{A}}_S} |\mathbf{x}_i^T \mathbf{d}|) = S(\tilde{\boldsymbol{\beta}}) + |\Delta y_l| \text{sign}(\Delta y_l) \tilde{\lambda}_q = S(\tilde{\boldsymbol{\beta}}) + \Delta y_l \tilde{\lambda}_q,$$

which completes the proof. ■

Remark 4.1 If the base matrix at $\tilde{\boldsymbol{\beta}}$ is ill-conditioned, then the change $\bar{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} = \Delta y_l \tilde{\mathbf{d}}_q$ may be much larger than Δy_l in magnitude, since $\|\tilde{\mathbf{d}}_q\|$ could be large. Similarly, the bound on $|\Delta y_l|$ may be small, since $|x_j^T \tilde{\mathbf{d}}_q|$ in (2.7) could be large. This highly sensitive behaviour is the same as occurs for perturbations to the responses y_i , as discussed in Shi and Lukas (2005).

4.2. Inactive inequality constraints

For a perturbation to an inactive inequality constraint, we have the following result.

Theorem 4.2 If $l \in \mathcal{I} \setminus \tilde{\mathcal{A}}_I$, then $\tilde{\boldsymbol{\beta}}$ is also an optimal base point for (1.1)^[l] provided $\Delta y_l \in [r_l(\tilde{\boldsymbol{\beta}}), \infty)$.

Proof Under the condition on Δy_l we have $r_l^{[l]}(\tilde{\boldsymbol{\beta}}) \equiv \mathbf{x}_l^T \tilde{\boldsymbol{\beta}} - (y_l + \Delta y_l) = r_l(\tilde{\boldsymbol{\beta}}) - \Delta y_l \leq 0$. Hence $\tilde{\boldsymbol{\beta}}$ is still feasible and therefore it is an optimal base point for (1.1)^[l]. ■

4.3. Active inequality constraints

For a perturbation to an active inequality constraint, the situation is similar to that for an equality constraint, as the next result shows.

Theorem 4.3 Let $l = j_q \in \tilde{\mathcal{A}}_I$ and let $\mathbf{d} = -\tilde{\mathbf{d}}_q$. If the perturbation Δy_l satisfies $|\Delta y_l| < \min\{\delta'(\mathbf{d}), \alpha'(\mathbf{d})\}$ (see (2.7)), then the new base point $\bar{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} - \Delta y_l \mathbf{d} = \tilde{\boldsymbol{\beta}} + \Delta y_l \tilde{\mathbf{d}}_q$ is a solution to (1.1)^[l] with the same optimal value as shown in (4.1).

Proof The proof is similar to that of Theorem 4.1. ■

5. Perturbations to the coefficients of constraints

Now we study the effect on the optimal solution of a perturbation to the gradient (coefficient vector) of a constraint. Suppose that $\tilde{\boldsymbol{\beta}}$ is the unique solution to the problem (1.1) and the k th constraint $\mathbf{x}_k^T \boldsymbol{\beta} - y_k = 0$ (or ≤ 0) is perturbed into $(\mathbf{x}_k + \Delta \mathbf{x}_k)^T \boldsymbol{\beta} - y_k = 0$ (or ≤ 0). The perturbed problem is denoted by (1.1)^{<k>}.

5.1. Equality constraints

When the gradient \mathbf{x}_k is active (i.e. for an equality constraint or active inequality constraint), the perturbed base matrix is not the same as the original one and we need to ensure that it is nonsingular. Clearly, for a sufficiently small change to a column vector, the nonsingularity of the matrix does not change. In fact we have the following result.

Proposition 5.1 Let the matrix $\bar{\mathbf{A}}$ be formed by replacing the q th column \mathbf{x}_k , $k = j_q$, of the nonsingular matrix $\tilde{\mathbf{A}}$ by $\mathbf{x}_k + \Delta \mathbf{x}_k$. Then $\bar{\mathbf{A}}$ is still nonsingular provided $\|\Delta \mathbf{x}_k\| < 1/\|\tilde{\mathbf{d}}_q\|$. Furthermore, the i th row of the inverse $\bar{\mathbf{D}} = \bar{\mathbf{A}}^{-1}$ is equal to the transpose of

$$\bar{\mathbf{d}}_i = \begin{cases} \frac{1}{1 + \Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_q} \tilde{\mathbf{d}}_q, & i = q, \\ \tilde{\mathbf{d}}_i - (\Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_i) \bar{\mathbf{d}}_q, & i \neq q. \end{cases} \quad (5.1)$$

See Shi and Lukas (2005) for the proof. The following result and its proof are similar to those for perturbations to the gradients of equations in Shi and Lukas (2005). But here the direction is not necessarily descent, and the optimality conditions have no limitation on the multipliers corresponding to equality constraints.

Theorem 5.1 Let $k = j_q \in \mathcal{E} \subset \tilde{\mathcal{A}}$ and suppose $r_k^{<k>}(\tilde{\boldsymbol{\beta}}) \equiv (\mathbf{x}_k + \Delta \mathbf{x}_k)^T \tilde{\boldsymbol{\beta}} - y_k = \Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}} \neq 0$. (In the case $\Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}} = 0$ including $\tilde{\boldsymbol{\beta}} = 0$, the point $\tilde{\boldsymbol{\beta}}$ remains the solution to the new problem). Let δ' and α' be the shortest steps to the boundary planes corresponding to the inactive equations and inactive constraints, respectively (see (2.7)), from $\tilde{\boldsymbol{\beta}}$ along the direction

$$\mathbf{d} = -\theta_k \tilde{\mathbf{d}}_q, \quad \theta_k \equiv \text{sign}(\Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}}). \quad (5.2)$$

If

$$\|\Delta \mathbf{x}_k\| < \eta \equiv \min\{\eta_0, \eta_1, \eta_2\}, \quad (5.3A)$$

$$\eta_0 \equiv \min\left\{\frac{1}{\|\tilde{\mathbf{d}}_q\|}, \frac{\delta'}{\|\tilde{\boldsymbol{\beta}}\| + \delta' \|\tilde{\mathbf{d}}_q\|}, \frac{\alpha'}{\|\tilde{\boldsymbol{\beta}}\| + \alpha' \|\tilde{\mathbf{d}}_q\|}\right\}, \quad (5.3B)$$

$$\eta_1 \equiv \min_{j_i \in \tilde{\mathcal{A}}_s} \frac{1 - |\tilde{\lambda}_i|}{|\tilde{\lambda}_q| \cdot \|\tilde{\mathbf{d}}_i\| + (1 - |\tilde{\lambda}_i|) \|\tilde{\mathbf{d}}_q\|}, \quad (5.3C)$$

$$\eta_2 \equiv \min_{j_i \in \tilde{\mathcal{A}}_I} \frac{|\tilde{\lambda}_i|}{|\tilde{\lambda}_q| \cdot \|\tilde{\mathbf{d}}_i\| + |\tilde{\lambda}_i| \cdot \|\tilde{\mathbf{d}}_q\|}, \quad (5.3D)$$

then

$$\bar{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} + \alpha^{<k>} \mathbf{d}, \quad \alpha^{<k>} = \frac{|\Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}}|}{1 + \Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_q}. \quad (5.4)$$

is the unique solution to (1.1)^{<k>} with the new optimal value

$$S^{<k>}(\bar{\boldsymbol{\beta}}) = S(\tilde{\boldsymbol{\beta}}) - \frac{(\Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}}) \tilde{\lambda}_q}{1 + \Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_q}.$$

Proof First we show that $\bar{\boldsymbol{\beta}}$, defined by (5.4), is feasible for (1.1)^{<k>}. For the new equality constraint, (5.4) and (5.2) imply that

$$\begin{aligned} r_k^{<k>}(\bar{\boldsymbol{\beta}}) &= (\mathbf{x}_k + \Delta \mathbf{x}_k)^T (\tilde{\boldsymbol{\beta}} - \theta_k \alpha^{<k>} \tilde{\mathbf{d}}_q) - y_k = r_k(\tilde{\boldsymbol{\beta}}) + \Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}} - \theta_k \alpha^{<k>} (\mathbf{x}_k + \Delta \mathbf{x}_k)^T \tilde{\mathbf{d}}_q \\ &= 0 + \Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}} - \theta_k \alpha^{<k>} (1 + \Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_q) = 0. \end{aligned}$$

For the other equality and inequality constraints at $\tilde{\boldsymbol{\beta}}$, we have

$$r_i^{<k>}(\bar{\boldsymbol{\beta}}) = \mathbf{x}_i^T (\tilde{\boldsymbol{\beta}} + \alpha^{<k>} \mathbf{d}) - y_i = r_i(\tilde{\boldsymbol{\beta}}) + \alpha^{<k>} \mathbf{x}_i^T \mathbf{d}. \quad (5.5)$$

For all $i \in \mathcal{E} \cup \tilde{\mathcal{A}}_I \setminus \{k\}$, it follows from (5.5) and $\mathbf{x}_i^T \mathbf{d} = 0$ that $r_i^{<k>}(\bar{\boldsymbol{\beta}}) = 0$. For all inactive (inequality) constraints at $\tilde{\boldsymbol{\beta}}$, from $\|\Delta \mathbf{x}_k\| < \eta_0$ in (5.3), we have $1 + \Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_q \geq 1 - |\Delta \mathbf{x}_k| \|\tilde{\mathbf{d}}_q\| \geq 1 - \|\Delta \mathbf{x}_k\| \cdot \|\tilde{\mathbf{d}}_q\| > 0$, and then

$$0 < \alpha^{<k>} \leq \|\Delta \mathbf{x}_k\| \cdot \|\tilde{\boldsymbol{\beta}}\| / (1 - \|\Delta \mathbf{x}_k\| \cdot \|\tilde{\mathbf{d}}_q\|) < \alpha'.$$

Hence, if $\mathbf{x}_i^T \mathbf{d} > 0$, then $r_i^{<k>}(\bar{\boldsymbol{\beta}}) < r_i(\tilde{\boldsymbol{\beta}}) + \alpha' \mathbf{x}_i^T \mathbf{d} \leq 0$ by (5.5) and (2.7), and, if $\mathbf{x}_i^T \mathbf{d} \leq 0$, then obviously $r_i^{<k>}(\bar{\boldsymbol{\beta}}) \leq r_i(\tilde{\boldsymbol{\beta}}) < 0$. Therefore $\bar{\boldsymbol{\beta}}$ is feasible for (1.1)^{<k>}. The above argument also shows that $\bar{\boldsymbol{\beta}}$ is a base point of (1.1)^{<k>} with the same active set as $\tilde{\boldsymbol{\beta}}$ for (1.1), since the perturbation does not affect any equation.

Now we show that $\bar{\boldsymbol{\beta}}$ is optimal for (1.1)^{<k>}. From above $\alpha^{<k>} < \alpha'$, and similarly we obtain $\alpha^{<k>} < \delta'$. Hence the move from $\tilde{\boldsymbol{\beta}}$ along \mathbf{d} with step length $\alpha^{<k>}$ does not cross any boundary planes associated with the inactive equations, so the “cost” vector for (1.1)^{<k>} is $\bar{\mathbf{c}} \equiv \bar{\mathbf{c}}(\bar{\boldsymbol{\beta}}) = \tilde{\mathbf{c}}$. But the base matrix at $\bar{\boldsymbol{\beta}}$ is \bar{A} as defined in Proposition 5.1, since $k = j_q \in \bar{A}$, the active set at $\bar{\boldsymbol{\beta}}$ for (1.1)^{<k>}. From (5.3) and Proposition 5.1, \bar{A} is nonsingular, and its inverse is given by (5.1). Consequently, the multipliers at $\bar{\boldsymbol{\beta}}$ for (1.1)^{<k>} are

$$\bar{\lambda}_i = \bar{\mathbf{d}}_i^T \bar{\mathbf{c}} = \bar{\mathbf{d}}_i^T \tilde{\mathbf{c}} = \begin{cases} \frac{\tilde{\lambda}_q}{1 + \Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_q}, & i = q \ (k = j_q), \\ \tilde{\lambda}_i - \frac{(\Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_i) \tilde{\lambda}_q}{1 + \Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_q}, & i \neq q. \end{cases}$$

Now we estimate the values of the multipliers. We do not need to check the value of $\tilde{\lambda}_q$ corresponding to an equality constraint. For the other multipliers, since $\|\Delta \mathbf{x}_k\| < \eta_1$ in (5.3),

$$|\bar{\lambda}_i| \leq |\tilde{\lambda}_i| + \frac{|\tilde{\lambda}_q| \cdot \|\tilde{\mathbf{d}}_i\| \cdot \|\Delta \mathbf{x}_k\|}{1 - \|\tilde{\mathbf{d}}_q\| \cdot \|\Delta \mathbf{x}_k\|} < 1, \quad j_i \in \tilde{\mathcal{A}}_S = \bar{\mathcal{A}}_S,$$

while, since $\|\Delta \mathbf{x}_k\| < \eta_2$,

$$\bar{\lambda}_i \leq -|\tilde{\lambda}_i| + \frac{|\tilde{\lambda}_q| \cdot \|\tilde{\mathbf{d}}_i\| \cdot \|\Delta \mathbf{x}_k\|}{1 - \|\tilde{\mathbf{d}}_q\| \cdot \|\Delta \mathbf{x}_k\|} < 0, \quad j_i \in \tilde{\mathcal{A}}_I = \bar{\mathcal{A}}_I.$$

Hence $\bar{\boldsymbol{\beta}}$ is the unique optimal point for (1.1)^{<k>}.

Finally, since the active set is unchanged and $\text{sign}(r_i^{<k>}(\bar{\boldsymbol{\beta}})) = \text{sign}(r_i(\tilde{\boldsymbol{\beta}})) \equiv \tilde{\sigma}_i$, $i \in \tilde{\mathcal{A}}_S$, the new optimal value is

$$\begin{aligned} S^{<k>}(\bar{\boldsymbol{\beta}}) &= \left(\sum_{i \in \tilde{\mathcal{A}}_S} + \sum_{i \in \tilde{\mathcal{A}}_S^c} \right) (|\mathbf{x}_i^T(\tilde{\boldsymbol{\beta}} - \theta_k \alpha^{<k>} \tilde{\mathbf{d}}_q) - y_i|) \\ &= \sum_{i \in \tilde{\mathcal{A}}_S} |0 - \theta_k \alpha^{<k>} \mathbf{x}_i^T \tilde{\mathbf{d}}_q| + \sum_{i \in \tilde{\mathcal{A}}_S^c} \tilde{\sigma}_i (\mathbf{x}_i^T \tilde{\boldsymbol{\beta}} - y_i) + \sum_{i \in \tilde{\mathcal{A}}_S^c} (-\theta_k \tilde{\sigma}_i \alpha^{<k>} \mathbf{x}_i^T \tilde{\mathbf{d}}_q) \\ &= 0 + S(\tilde{\boldsymbol{\beta}}) - \theta_k \alpha^{<k>} \tilde{\mathbf{c}}^T \tilde{\mathbf{d}}_q = S(\tilde{\boldsymbol{\beta}}) - \frac{(\Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}}) \tilde{\lambda}_q}{1 + \Delta \mathbf{x}_k^T \tilde{\mathbf{d}}_q} \end{aligned}$$

which completes the proof. ■

Remark 5.1 This result indicates there is potentially high sensitivity if either the base matrix at $\tilde{\boldsymbol{\beta}}$ is ill-conditioned or if $\tilde{\boldsymbol{\beta}}$ is close to being non-unique (with $|\tilde{\lambda}_i| \approx 1$, $j_i \in \tilde{\mathcal{A}}_S$, or $\tilde{\lambda}_i \approx 0$, $j_i \in \tilde{\mathcal{A}}_I$). In these cases the bound on $\|\Delta \mathbf{x}_k\|$ in (5.3) may be small while $\|\tilde{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}\|$ could be much larger. These highly sensitive cases are the same as those for perturbations to the row vectors \mathbf{x}_i^T of the design matrix, as discussed in Shi and Lukas (2005).

5.2. Inactive inequality constraints

Theorem 5.2 Let $k \in \mathcal{I} \setminus \tilde{\mathcal{A}}_I$ so $r_k(\tilde{\boldsymbol{\beta}}) < 0$. Then $\tilde{\boldsymbol{\beta}}$ is also an optimal base point for (1.1)^{<k>} provided $\Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}} \leq -r_k(\tilde{\boldsymbol{\beta}})$. **Proof** With the condition on $\Delta \mathbf{x}_k$, we have $r_k^{<k>}(\tilde{\boldsymbol{\beta}}) = r_k(\tilde{\boldsymbol{\beta}}) + \Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}} \leq 0$. Hence $\tilde{\boldsymbol{\beta}}$ is still feasible and the result follows. ■

5.3. Active inequality constraints

Theorem 5.3 Let $k = j_q \in \tilde{\mathcal{A}}_I$ and suppose $r_k^{<k>}(\tilde{\boldsymbol{\beta}}) = \Delta \mathbf{x}_k^T \tilde{\boldsymbol{\beta}} \neq 0$. If $\|\Delta \mathbf{x}_k\|$ satisfies (5.3), then the conclusion of Theorem 5.1 is also valid here.

Proof The proof of this case is similar to that of Theorem 5.1. In fact, regarding this active inequality constraint as an equality constraint of (1.1) does not affect the solution of (1.1). So, with the same arguments as those for Theorem 5.1, we derive the same conclusion. ■

6. Deletion of an observation

In assessing an L_1 regression estimate, the effect of deleting each observation provides very useful diagnostic information. It identifies the most influential observations and others that have little or no effect on the solution. The effect of deleting one or more observations is also central in cross-validation for model selection, and in jackknife estimation of variance (for L_1 regression, a delete- d jackknife should be used since the estimator is not smooth (see Effron and Tibshirani (1993), Sec. 11.6 and 11.7)).

In this section, let $(1.1)^{\{-t\}}$ denote the problem obtained by deleting the t -th observation from (1.1). Let $\tilde{\boldsymbol{\beta}}$ be an optimal base point for (1.1) (not necessarily unique).

Since deletion (and addition) of an observation is a discrete change to the problem, there can be a sizeable change in the optimal solution. There may be no change to the active set, a change of one active equation or constraint or perhaps even several. For this reason it is hard to find a general estimate or bound on the change in the solution. However, it is possible to identify cases for which there will be no change in the solution. Also, in general, one can use the RGA algorithm starting at $\tilde{\boldsymbol{\beta}}$ to find the new solution, which is far more efficient than solving the new problem from the beginning.

6.1. Inactive observation

First we consider the case where the deleted observation (equation) is inactive.

Theorem 6.1 If $t \in \tilde{\mathcal{A}}_S^c$, then $\tilde{\boldsymbol{\beta}}$ remains a solution of $(1.1)^{\{-t\}}$ if and only if

$$\begin{cases} |\tilde{\lambda}_i - \sigma_t \tilde{\mathbf{d}}_i^T \mathbf{x}_t| \leq 1, & \text{if } j_i \in \tilde{\mathcal{A}}_S, \\ \tilde{\lambda}_i - \sigma_t \tilde{\mathbf{d}}_i^T \mathbf{x}_t \leq 0, & \text{if } j_i \in \tilde{\mathcal{A}}_I, \\ \tilde{\lambda}_i - \sigma_t \tilde{\mathbf{d}}_i^T \mathbf{x}_t = 0, & \text{if } j_i \in \tilde{\mathcal{A}}_0, \end{cases} \quad (6.1)$$

where $\sigma_t = \text{sign}(r_t(\tilde{\boldsymbol{\beta}}))$.

Proof In this case, from (2.3A) the new “cost” vector is

$$\mathbf{c}^{\{-t\}} = \tilde{\mathbf{c}} - \sigma_t \mathbf{x}_t, \quad (6.2)$$

and the inverse of the base matrix is not changed. Hence the new multipliers for $(1.1)^{\{-t\}}$ at $\tilde{\boldsymbol{\beta}}$ are

$$\lambda_i^{\{-t\}} = \tilde{\mathbf{d}}_i^T \mathbf{c}^{\{-t\}} = \tilde{\mathbf{d}}_i^T \tilde{\mathbf{c}} - \sigma_t \tilde{\mathbf{d}}_i^T \mathbf{x}_t = \tilde{\lambda}_i - \sigma_t \tilde{\mathbf{d}}_i^T \mathbf{x}_t. \quad (6.3)$$

Comparing (6.3) with (2.5) shows that (6.1) is a necessary and sufficient condition for $\tilde{\boldsymbol{\beta}}$ to remain a solution of $(1.1)^{\{-t\}}$. ■

6.2. Active observation

If the deleted equation is active, we have the following.

Theorem 6.2 Suppose $t = j_q \in \tilde{\mathcal{A}}_S$. Then $\tilde{\boldsymbol{\beta}}$ remains a solution for $(1.1)^{\{-t\}}$ if and only if the corresponding multiplier $\tilde{\lambda}_q = 0$. If $\tilde{\lambda}_q \neq 0$, the direction

$$\mathbf{d} = -\text{sign}(\tilde{\lambda}_q) \tilde{\mathbf{d}}_q \quad (6.4)$$

is a feasible descent direction. Furthermore, if index p is the new element of the active set $\mathcal{A}^{\{-t\}}$ at the new (feasible) base point $\tilde{\boldsymbol{\beta}}^{\{-t\}} = \tilde{\boldsymbol{\beta}} + \alpha \mathbf{d}$ obtained by a line search from $\tilde{\boldsymbol{\beta}}$ along \mathbf{d} (with α chosen as in step (4) of the RGA in Section 2), the new multipliers are

$$\bar{\lambda}_i^{\{-t\}} = \begin{cases} \tilde{\lambda}_i / \mathbf{x}_p^T \tilde{\mathbf{d}}_q, & i = q, \\ \tilde{\lambda}_i + (\mathbf{x}_p^T \tilde{\mathbf{d}}_i / \mathbf{x}_p^T \tilde{\mathbf{d}}_q) \tilde{\lambda}_q, & i \neq q. \end{cases} \quad (6.5)$$

So $\tilde{\boldsymbol{\beta}}^{\{-t\}}$ is a solution to $(1.1)^{\{-t\}}$ provided all $\lambda_i^{\{-t\}}$ satisfy (2.5).

Proof In this case, we associate the index t with an artificial constraint $\mathbf{x}_t^T \boldsymbol{\beta} - y_t \leq 0$ for $(1.1)^{\{-t\}}$, which is active at $\tilde{\boldsymbol{\beta}}$. With this point of view, $\tilde{\boldsymbol{\beta}}$ is also a base point of $(1.1)^{\{-t\}}$ with the same inverse base matrix \tilde{D} , and the “cost” vector of $(1.1)^{\{-t\}}$, $\mathbf{c}^{\{-t\}}$, is the same as the original one, i.e. $\mathbf{c}^{\{-t\}} = \tilde{\mathbf{c}}$. Hence, $\lambda_i^{\{-t\}} = \tilde{\lambda}_i$, $i = 1, \dots, n$. Since the t -th equation for $(1.1)^{\{-t\}}$ is artificial, $\tilde{\boldsymbol{\beta}}$ remains a solution for $(1.1)^{\{-t\}}$ if and only if the corresponding multiplier $\lambda_q^{\{-t\}} = \tilde{\lambda}_q = 0$.

If $\tilde{\lambda}_q \neq 0$, the direction defined in (6.4) is a feasible descent direction according to (2.6). Furthermore, from (2.4) and (2.8), the new multipliers are defined by (6.5) and the result follows. ■

Example 6.1 (Shi (1997) Example 2.4.1 and Example 7.2.1) Consider the LL_1 problem:

$$\begin{aligned} \text{minimize} \quad & S(\boldsymbol{\beta}) = |2\beta_1 + \beta_2 - 4| + |-\beta_1 + \beta_2 - 1| + |\beta_1 - 3| + |\beta_1 + \beta_2 - 2| + |\beta_1 - 4| \\ \text{subject to} \quad & -\beta_1 + 2\beta_2 - 2 \leq 0, \quad -\frac{5}{3}\beta_1 - \beta_2 \leq 0, \quad \beta_1 \geq -1, \quad \beta_2 \geq -2.5. \end{aligned}$$

It is easy to check that an optimal solution, its active set and multipliers are $\tilde{\boldsymbol{\beta}} = (2, 0)^T$, $\tilde{\mathcal{A}} = \{1, 4\}$ and $(\tilde{\lambda}_1, \tilde{\lambda}_2) = (0, -1)$, respectively. The point $\tilde{\boldsymbol{\beta}}$ remains optimal for $(1.1)^{\{-1\}}$ since $\lambda_1^{\{-1\}} = \tilde{\lambda}_1 = 0$, but it is not a solution of $(1.1)^{\{-4\}}$ since $\lambda_2^{\{-4\}} = \tilde{\lambda}_2 = -1 \neq 0$. It is easy to solve the latter problem from the beginning using the RGA. However, it is more efficient to start from $\tilde{\boldsymbol{\beta}}$. From (6.4), a descent direction at $\tilde{\boldsymbol{\beta}}$ for $(1.1)^{\{-4\}}$ is $\mathbf{d} = \tilde{\mathbf{d}}_2 = (-1, 2)^T$. The search along the edge of P_1 will reach the optimal point $\boldsymbol{\beta}^{\{-4\}} = (1.2, 1.6)^T$ with active set $\mathcal{A}^{\{-4\}} = \{1, 6\}$ and multipliers $\lambda_1^{\{-4\}} = -3/5$, $\lambda_2^{\{-4\}} = -1/5$.

Remark 6.1 This example and Theorems 6.1 and 6.2 show that if $\tilde{\boldsymbol{\beta}}$ is not a solution of $(1.1)^{\{-t\}}$, we can use $\tilde{\boldsymbol{\beta}}$ as a starting base point for the RGA with the feasible descent direction defined by (6.4) if $t \in \tilde{\mathcal{A}}_S$ and by (2.6) and (6.3) if $t \in \tilde{\mathcal{A}}_S^c$. This is more efficient than solving $(1.1)^{\{-t\}}$ from the very beginning.

For the model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where X is $n \times p$, the influence of the t -th observation on a regression estimate $\tilde{\boldsymbol{\beta}}$ can be measured by $V_t \equiv \|X\tilde{\boldsymbol{\beta}} - X\tilde{\boldsymbol{\beta}}^{\{-t\}}\|$, where $\tilde{\boldsymbol{\beta}}^{\{-t\}}$ is the regression estimate obtained with the t -th observation deleted. Apart from a scale factor, V_t is the Cook distance, which was introduced for least squares regression by Cook (1977) (see also Cook and Weisberg (1982)). In the least squares case, there is a formula for V_t , which means one doesn't have to solve n additional regression problems to find V_t , $t = 1, \dots, n$.

For L_1 regression, although there is no simple formula for V_t , it can still be computed efficiently by computing $\tilde{\boldsymbol{\beta}}^{\{-t\}}$ using $\tilde{\boldsymbol{\beta}}$ as a starting point. For an unconstrained L_1 regression problem, Castillo et al. (2001) computed V_t using the LP formulation and a variant of the dual simplex method (called the Exterior Point Method) to compute $\tilde{\boldsymbol{\beta}}^{\{-t\}}$ starting from $\tilde{\boldsymbol{\beta}}$. For the general LL_1 problem (1.1), it is clear from Remark 6.1 that the RGA can be used to efficiently compute V_t .

7. Addition of an observation

Suppose we have a new observation $y_t = \mathbf{x}_t^T \boldsymbol{\beta} + \varepsilon_t$, and the term $|\mathbf{x}_t^T \boldsymbol{\beta} - y_t|$ is added to $S(\boldsymbol{\beta})$ in (1.1A) forming a new problem $(1.1)^{\{+t\}}$. Let $r_t(\boldsymbol{\beta}) = \mathbf{x}_t^T \boldsymbol{\beta} - y_t$ be the new residual.

First we show that if the new observation has a certain property then an infinite solution set becomes smaller.

Theorem 7.1 Let \mathcal{M} be the solution set of the original problem and suppose the new observation divides \mathcal{M} , i.e. there are two points $\tilde{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\beta}}$ of \mathcal{M} such that $r_t(\tilde{\boldsymbol{\beta}})$ and $r_t(\bar{\boldsymbol{\beta}})$ are of opposite sign, then the solution set of $(1.1)^{\{+t\}}$ is $\mathcal{M} \cap P_t$, where $P_t = \{\boldsymbol{\beta} \mid \mathbf{x}_t^T \boldsymbol{\beta} - y_t = 0\}$ is the new boundary plane.

Proof Let $S^{\{+t\}}(\boldsymbol{\beta}) = S(\boldsymbol{\beta}) + |r_t(\boldsymbol{\beta})|$ be the new sum of absolute deviations at $\boldsymbol{\beta}$. Clearly, for any $\tilde{\boldsymbol{\beta}} \in \mathcal{M} \setminus P_t$, there is a $\bar{\boldsymbol{\beta}} \in \mathcal{M} \setminus P_t$ for which the corresponding residuals have opposite signs. Since \mathcal{M} is convex, the line between $\tilde{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\beta}}$ lies in \mathcal{M} and intersects P_t at some point $\boldsymbol{\beta}^*$, and $S^{\{+t\}}(\boldsymbol{\beta}^*) = S(\boldsymbol{\beta}^*) = S(\tilde{\boldsymbol{\beta}}) < S^{\{+t\}}(\tilde{\boldsymbol{\beta}})$. Also, for any $\boldsymbol{\beta} \notin \mathcal{M}$, we have $S^{\{+t\}}(\boldsymbol{\beta}^*) = S(\boldsymbol{\beta}^*) < S(\boldsymbol{\beta}) \leq S^{\{+t\}}(\boldsymbol{\beta})$. Since $S^{\{+t\}}(\boldsymbol{\beta}) = S^{\{+t\}}(\boldsymbol{\beta}^*)$ for all $\boldsymbol{\beta} \in \mathcal{M} \cap P_t$, the result follows. ■

Now we derive conditions under which a solution remains optimal. Let $\tilde{\boldsymbol{\beta}}$ be an optimal base point for (1.1) (not necessarily unique).

Theorem 7.2 Suppose we have a new observation $y_t = \mathbf{x}_t^T \boldsymbol{\beta} + \varepsilon_t$, giving a new problem (1.1) $^{\{+t\}}$. If $r_t(\tilde{\boldsymbol{\beta}}) = 0$, then $\tilde{\boldsymbol{\beta}}$ remains a solution of (1.1) $^{\{+t\}}$. If $r_t(\tilde{\boldsymbol{\beta}}) \neq 0$, with $\sigma_t = \text{sign}(r_t(\tilde{\boldsymbol{\beta}}))$, then $\tilde{\boldsymbol{\beta}}$ remains a solution of (1.1) $^{\{+t\}}$ if and only if

$$\begin{cases} |\tilde{\lambda}_i + \sigma_t \tilde{\mathbf{d}}_i^T \mathbf{x}_t| \leq 1, & \text{if } j_i \in \tilde{\mathcal{A}}_S, \\ \tilde{\lambda}_i + \sigma_t \tilde{\mathbf{d}}_i^T \mathbf{x}_t \leq 0, & \text{if } j_i \in \tilde{\mathcal{A}}_I, \\ \tilde{\lambda}_i + \sigma_t \tilde{\mathbf{d}}_i^T \mathbf{x}_t = 0, & \text{if } j_i \in \tilde{\mathcal{A}}_0. \end{cases} \quad (7.1)$$

Proof In this case $t \in \tilde{\mathcal{A}}_S^c$ and the new “cost” vector is $\mathbf{c}^{\{+t\}} = \tilde{\mathbf{c}} + \sigma_t \mathbf{x}_t$, which is different from $\mathbf{c}^{\{-t\}}$ in (6.2) only in the middle operation. The result follows by the same argument as for Theorem 6.1. ■

This result demonstrates the robustness of L_1 regression with respect to additional data values, since the condition (7.1) depends on y_t only through the sign of the residual $r_t(\tilde{\boldsymbol{\beta}})$. As a special case, consider an unconstrained problem with solution $\tilde{\boldsymbol{\beta}}$ and a new observation $y_t = \mathbf{x}_t^T \boldsymbol{\beta} + \varepsilon_t$, where \mathbf{x}_t is the same as an active gradient vector \mathbf{x}_{j_q} , $j_q \in \tilde{\mathcal{A}}$. Then, since $\tilde{\mathbf{d}}_i^T \mathbf{x}_t = \delta_{iq}$, the condition (7.1) simplifies to $|\tilde{\lambda}_q + \sigma_t| \leq 1$. Hence $\tilde{\boldsymbol{\beta}}$ remains a solution if and only if $\tilde{\lambda}_q = 0$ or $\tilde{\lambda}_q$ and $r_t(\tilde{\boldsymbol{\beta}})$ have opposite signs (independent of the magnitude of y_t).

Example 7.1 Consider $S(\boldsymbol{\beta})$ in Example 3.1 and add the term $|2\beta_1 + \beta_2 - y_t|$. Then $\mathbf{x}_t = \mathbf{x}_1$, which is active for both optimal base points $(2, 0)^T$ and $(1, 2)^T$ in Example 3.1. For $(2, 0)^T$ the corresponding multiplier is $\tilde{\lambda}_1 = 0$ so $(2, 0)^T$ remains a solution of the new problem for arbitrary values of y_t . The same is true for $(1, 2)^T$, and so the whole solution set remains optimal for the new problem. (In fact, the new problem has a larger solution set, unlike the situation in Theorem 7.1.)

8. Numerical results

Example 8.1 Derivative estimation under a concavity constraint

We will consider the derivative estimation problem discussed in Shi and Lukas (2005). We wish to estimate the derivative of a function $g(t)$, $0 \leq t \leq 1$, satisfying $g(0) = 0$, given discrete noisy data $y_i = g(t_i) + \varepsilon_i$, $i = 1, \dots, n$. This problem arises in many applications, including the analysis of growth curves (Gasser et al. (1984) and Eubank (1988)) and pharmacokinetic response curves (Song et al. (1995)). The estimated derivative can be used to determine characteristic time points like the point of maximum growth rate or the point of peak response.

Numerical differentiation is an ill-posed problem (see Anderssen and Bloomfield (1974)) and requires some stabilization to obtain reasonable solutions. A simple approach to this is to approximate $g'(t)$ in a low dimensional subspace (by some convenient parametric form), where the dimension plays the role of a smoothing parameter.

Clearly the true solution $f(t) = g'(t)$ satisfies $\int_0^t f(\bar{t}) d\bar{t} = g(t)$, since $g(0) = 0$. As in Shi and Lukas (2005) we approximate $f(t)$ by the truncated trigonometric series

$$f_p(t) = c_0 + \sum_{k=1}^{p_1} c_k \cos 2k\pi t + \sum_{k=1}^{p_2} d_k \sin 2k\pi t, \quad (8.1)$$

where $p = 1 + p_1 + p_2 \ll n$, and evaluate the indefinite integral at the points t_i , giving the overdetermined system

$$\int_0^{t_i} f_p(\bar{t}) d\bar{t} = c_0 t_i + \sum_{k=1}^{p_1} c_k \sin 2k\pi t_i / (2k\pi) + \sum_{k=1}^{p_2} d_k (1 - \cos 2k\pi t_i) / (2k\pi) \approx y_i, \quad i = 1, \dots, n. \quad (8.2)$$

Applying L_1 regression, we have a problem of the form of (1.1A) with $\beta = (c_0, c_1, \dots, c_{p_1}, d_1, \dots, d_{p_2})^T$.

In many situations there is additional information about the unknown true derivative. In the case of child growth curves, obviously the growth rate (derivative) is non-negative for all time. In the case of pharmacokinetic response curves, the underlying reaction kinetics model determines properties of the reaction rate $r = dc/dt$; e.g., typically the (negative) rate increases monotonically after a certain time, so $dr/dt > 0$. It is worth trying to incorporate such extra information into the estimation procedure above.

For our simulations we will assume it is known that the true derivative $f(t) = g'(t)$ is concave, i.e. $f''(t) \leq 0$ for all t . A concrete example is the estimation of the speed of a vehicle from time and position data, given that the acceleration is decreasing. To use this information, we evaluate $f_p''(t) \leq 0$ at finitely many values of t , resulting in constraints of the form in (1.1C). Thus we obtain a constrained L_1 regression problem.

For the numerical experiments, we take $g(t) = 10(-t^3/3 + t^2/2)$ and so $g'(t) = 10(-t^2 + t)$. The data are $y_i = g(t_i) + \varepsilon_i$, $i = 1, \dots, n$, where $n = 50$, $t_i = i/n$ and the ε_i are pseudo-random numbers distributed normally with mean 0 and standard deviation 0.01, i.e. approximately 1% noise. The L_1 estimate of $g'(t)$ is computed by solving the constrained L_1 regression problem defined above with $p_1 = p_2 = 5$ so $p = 11$. The computations were done in MATLAB, using the linprog routine from the Optimization Toolbox to solve the LP formulation of the L_1 regression problem as given in Barrodale and Roberts (1978). The results were confirmed using the Fortran 77 routine for the RGA described in Shi and Lukas (2002).

The function $g(t)$ and data are shown in Fig. 1(a). Fig. 1(b) shows the derivative $g'(t)$ and the unconstrained L_1 regression estimate of $g'(t)$.

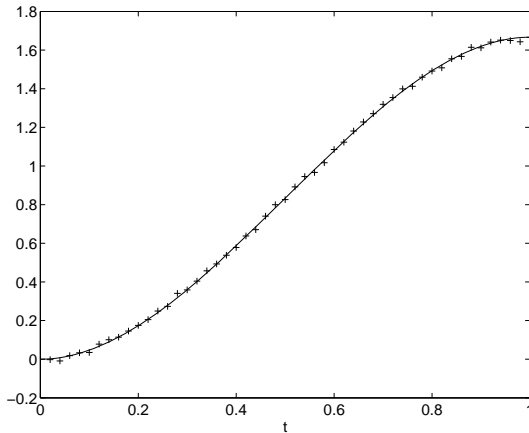


Fig. 1(a) $g(t) = 10(-t^3/3 + t^2/2)$ (solid),
and data (+)

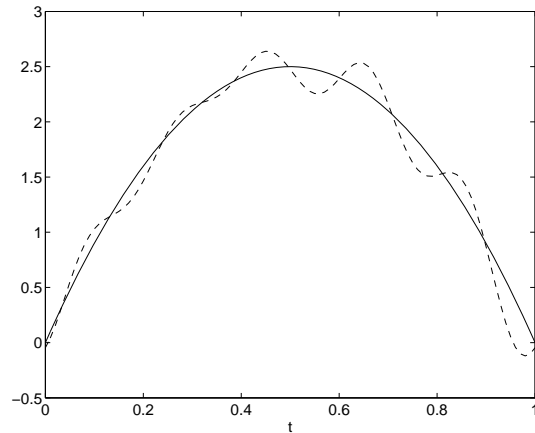


Fig. 1(b) $g'(t) = 10(-t^2 + t)$ (solid), and
unconstrained L_1 regression estimate (dashed)

Now we use the information that $f(t) = g'(t)$ is concave. With 5 constraints $f_p''(T_i) \leq 0$, where $T_i = 0.4 + 0.05(i-1)$, $i = 1, 2, \dots, 5$, we obtain the L_1 estimate of $g'(t)$ shown in Fig. 2(a). Clearly, the constrained estimate is better than the unconstrained one in Fig. 1(b). For the constrained estimate, two of the five constraints are active, i.e. $f_p''(0.4) = 0$ and $f_p''(0.55) = 0$, as shown by the two + marks on the graph of $f_p''(t)$ in Fig. 2(b).

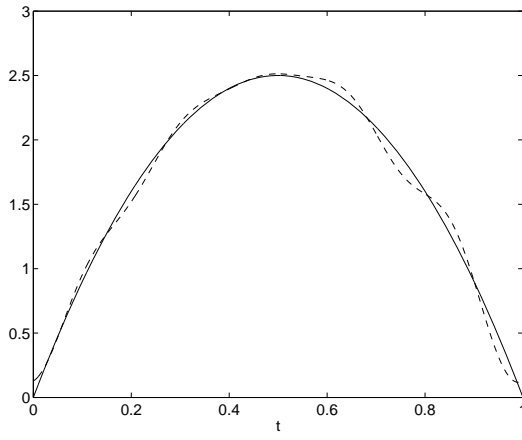


Fig. 2(a) $g'(t) = 10(-t^2 + t)$ (solid), and
constrained L_1 regression estimate $f_p(t)$ (dashed)

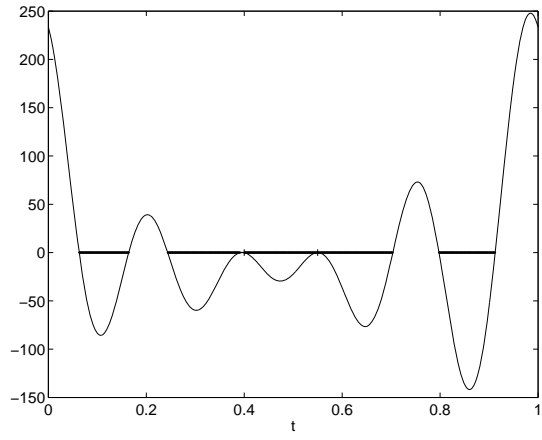


Fig. 2(b) $f_p''(t)$ for constraint
points 0.4, 0.45, \dots , 0.6

Let $\tilde{\boldsymbol{\beta}}$ be the optimal solution of the constrained L_1 regression problem defining the coefficients of $f_p(t)$. In the following subsections we will consider different types of perturbations to this L_1 regression problem, possibly resulting in a new optimal solution $\bar{\boldsymbol{\beta}}$. If $\tilde{f}_p(t)$ and $\bar{f}_p(t)$ are the estimates of $g'(t)$ corresponding to $\tilde{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\beta}}$, respectively, then it is not hard to show from (8.1) that $\|\tilde{f}_p - \bar{f}_p\|_\infty \leq \|\tilde{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}\|_1$. Therefore, a bound on $\|\tilde{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}\|_1$ will also serve as a bound on the change in the estimate of $g'(t)$.

8.1. Perturbations to responses

First we investigate the sensitivity of the constrained L_1 solution $\tilde{\boldsymbol{\beta}}$ for perturbations to the response vector \mathbf{y} . This case is covered by Theorem 3.1 of Shi and Lukas (2005). In simplified form, this states that for a sufficiently small perturbation vector $\Delta\mathbf{y}$, the solution $\bar{\boldsymbol{\beta}}$ of the perturbed problem satisfies

$$\|\bar{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\|_1 \leq p \|\tilde{D}\|_\infty \|\Delta\mathbf{y}\|_\infty. \quad (8.3)$$

This result was illustrated in Shi and Lukas (2005) by simulations with the unconstrained L_1 regression estimate of $g'(t)$. Here we illustrate it with the constrained L_1 regression estimate. As in Shi and Lukas (2005), the simulations involved taking 40 random perturbation vectors $\Delta\mathbf{y}$ with components $\Delta y_i = \pm 10^{-13}$, where the sign is pseudo-random with probability 0.5. To see the effect of the original data set, this was repeated 10 times with different data vectors \mathbf{y} generated by taking different sets of pseudo-random errors ε_i (all normally distributed with mean 0 and standard deviation 0.01). The size of the perturbations was chosen to be 10^{-13} to ensure that the conditions of Theorem 3.1 of Shi and Lukas (2005) hold for all the problems in the simulation. The resulting L_1 norm $\|\bar{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\|_1$ of errors in the solution and the bound in (8.3) are plotted in Fig. 3. Clearly, (as in the unconstrained case in Shi and Lukas (2005)) the bound is quite good and follows the behaviour of the errors over the 10 replicates.

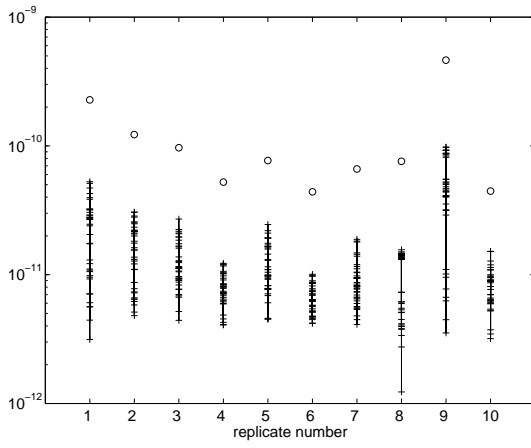


Fig. 3 L_1 norm of error in coefficients $\beta(+)$ and bound (o) for constrained L_1 estimates of $g'(t)$

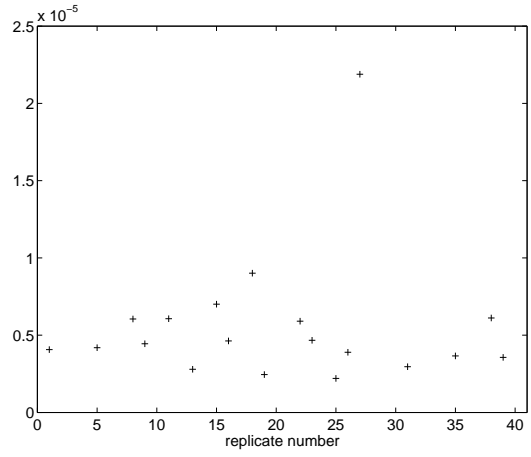


Fig. 4 L_1 norm of error in coefficients $\beta(+)$ for perturbation $f''(0.55) \leq -0.5$

8.2. Perturbations to constants of constraints

For the optimal solution $\tilde{\beta}$ to the constrained L_1 regression problem above there are 2 active constraints, corresponding to $f_p''(0.4) = 0$ and $f_p''(0.55) = 0$. To investigate the sensitivity of this solution and illustrate Theorem 4.3, we perturb one inequality constraint to become $f_p''(0.55) \leq -0.001$ and compute the new solution $\bar{\beta}$ and the error $\|\bar{\beta} - \tilde{\beta}\|_1$. This is repeated for 40 replicates of the data vector \mathbf{y} , and the resulting errors (for the 19 replicates in which this constraint was active) are plotted in Fig. 4. Clearly, the solution is not sensitive to this perturbation. These results are consistent with Theorem 4.3; for each replicate the perturbation $\Delta y_l = -0.001$ satisfies the condition in Theorem 4.3 and so the error is simply $\|\bar{\beta} - \tilde{\beta}\|_1 = 0.001\|d_q\|_1$, where $l = j_q$.

8.3. Perturbations to coefficients of constraints

For the optimal solution $\tilde{\beta}$ to the constrained L_1 regression problem above, 3 of the 5 constraints are inactive. In particular, the last constraint, i.e. $f_p''(0.6) \leq 0$, is inactive, with the left hand side equal to $r_k(\tilde{\beta}) = \mathbf{x}_k^T \tilde{\beta} - y_k = -36.448$, where $k = 55$.

From Theorem 5.2, for a perturbation $\Delta \mathbf{x}_k^T$ of \mathbf{x}_k^T in an inactive constraint, the solution $\tilde{\beta}$ remains optimal if the perturbation satisfies $r_k^{<k>}(\tilde{\beta}) = r_k(\tilde{\beta}) + \Delta \mathbf{x}_k^T \tilde{\beta} \leq 0$. We will consider the important perturbation resulting from a change in a constraint point, in particular from $T_5 = 6$ to \bar{T}_5 . For this it is easiest to find $r_k^{<k>}(\tilde{\beta})$, $k = 55$, directly as $r_k^{<k>}(\tilde{\beta}) = f_p''(\bar{T}_5)$. From the graph of $f_p''(t)$ in Fig. 2(b), it is clear that the solution $\tilde{\beta}$ remains optimal if \bar{T}_5 belongs to the set $\mathcal{V}_5 = [0.063, 0.164] \cup [0.244, 0.391] \cup [0.4, 0.704] \cup [0.798, 0.912]$ (to 3 decimal places), which is marked in Fig. 2(b) by a thick line.

Clearly, with the constraint points $0.4, 0.45, \dots, 0.6$, the L_1 regression estimate $f_p(t)$ satisfies the desired condition $f_p''(t) \leq 0$ for all $t \in \mathcal{V}_5$. The set \mathcal{V}_5 covers much of the interval $[0, 1]$ (the length of \mathcal{V}_5 is 0.667) but not close to all of it. This raises the design question of where to place additional constraint points so that the condition is satisfied over a larger set. It is clear from above that to change the solution, any shifted or new constraint point must be outside \mathcal{V}_5 .

To investigate the effect of additional constraints, we expanded the set of constraints to $f_p''(T_i) \leq 0$, $T_i = 0.1 + 0.05(i - 1)$, $i = 1, \dots, 17$, and recomputed the constrained L_1 regression solution. The new L_1 regression estimate $f_p(t)$, shown in Fig. 5(a), appears to be a better fit than the estimate in Fig. 2(a) based on 5 constraints. From Fig. 5(b) it is clear that for the new estimate, the second derivative satisfies

$f_p''(t) \leq 0$ on a larger set \mathcal{V}_{17} (with length 0.715) than \mathcal{V}_5 in Fig. 2(b). Now there are 4 active constraints corresponding to the constraint points 0.25, 0.3, 0.75 and 0.8, marked by + signs in Fig. 5(b).

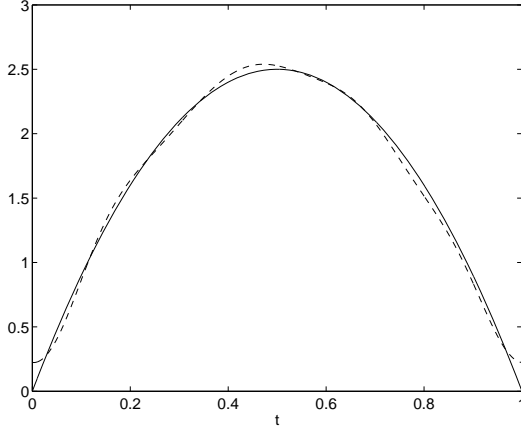


Fig. 5(a) $g'(t) = 10(-t^2 + t)$ (solid), and constrained L_1 estimate $f_p(t)$ (dashed)

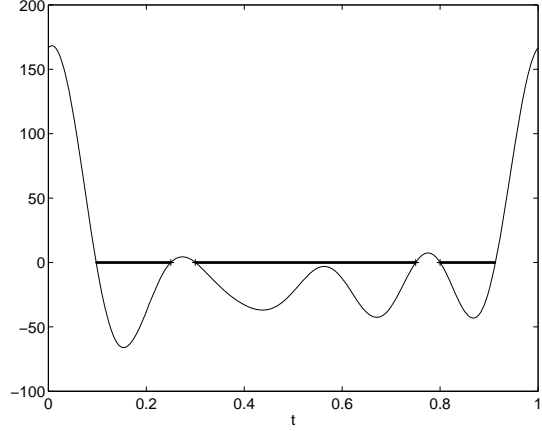


Fig. 5(b) $f_p''(t)$ for constraint points 0.1, 0.15, ..., 0.9

The computations above were replicated 10 times using 10 different data sets $y_i = g(t_i) + \varepsilon_i$, $i = 1, \dots, n$, with different pseudo-random errors (but with the same mean 0 and standard deviation 0.01). For each data set, we computed the constrained L_1 regression estimate $f_p(t)$ using the constraint points 0.4, 0.45, ..., 0.6 and the estimate using the constraint points 0.1, 0.15, ..., 0.9. For the two estimates we computed the lengths of the sets \mathcal{V}_5 and \mathcal{V}_{17} of points t such that $f_p''(t) \leq 0$, and these lengths are plotted in Fig. 6(a). Clearly, the length of \mathcal{V}_{17} is significantly greater than that of \mathcal{V}_5 for all the replicates. The estimates themselves were also compared using the following norm (a discrete L_1 Sobolev norm) of the error:

$$\|f_p - f\|_W = \frac{1}{1001} \left(\sum_{i=0}^{1000} |f_p(i/1000) - f(i/1000)| + \sum_{i=0}^{1000} |f_p''(i/1000) - f''(i/1000)| \right),$$

where $f(t) = 10(-t^2 + t)$ and $f''(t) = -20$. Note that this error norm takes into account not only deviations in function values but also deviations in curvature as determined by the second derivative. For each of the 10 replicates, the error norms for the two constrained estimates are plotted in Fig. 6(b). Clearly, for all the replicates, the estimate using the constraint points 0.1, 0.15, ..., 0.9 is significantly more accurate with respect to this norm.

It is clear from Fig. 5(b) that the optimal solution with the 17 constraint points 0.1, 0.15, ..., 0.9 is the same as the optimal solution with the four constraint points 0.25, 0.3, 0.75, 0.8 (the ones corresponding to active constraints). In fact, for all ten replicates, the optimal solution with the 17 constraint points is the same as the optimal solution with a smaller set of (different) constraint points, of sizes 4, 6, 5, 5, 4, 5, 3, 4, 5 and 3 points (i.e. the number of active constraints).

If the set of constraint points is expanded to $\{0.05, 0.1, 0.15, \dots, 0.9, 0.95\}$ (i.e. two extra points), there is a dramatic change in the optimal solution in that it becomes a constant $f_p(t) = 1.784$, which is obviously a poor estimate of $g'(t)$. Essentially, the constraints have now completely knocked out the sinusoidal terms from the estimate $f_p(t)$ in (8.1), leaving only the term $c_0 t$ in (8.2) to fit to the data by least absolute deviations. For this constant solution all the constraints are active. It is clear that this problem would not

occur if the number p of terms in $f_p(t)$ were increased sufficiently (i.e. increasing the degrees of freedom). In practice, however, p would usually be chosen first and then the constraint points would be decided. The example shows that some care should be taken in the selection of the constraint points.

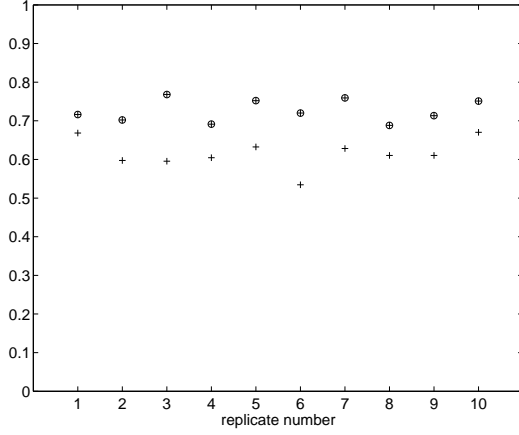


Fig. 6(a) Lengths of $\{t : f_p''(t) \leq 0\}$ for
0.4, 0.45, \dots , 0.6 (+) and 0.1, 0.15, \dots , 0.9 (\oplus)

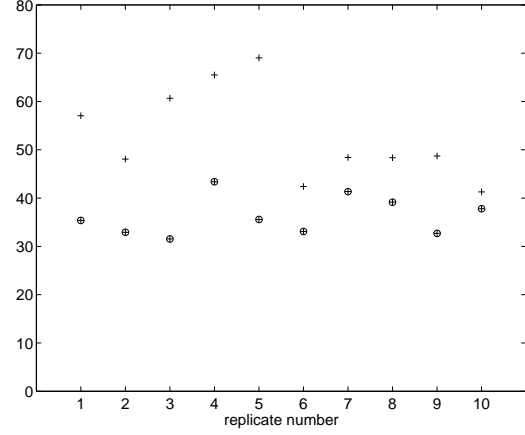


Fig. 6(b) Errors $\|f_p - f\|_W$ for constraint points
0.4, 0.45, \dots , 0.6 (+) and 0.1, 0.15, \dots , 0.9 (\oplus)

8.4. Deletion of observations

Lastly we investigate the influence of each observation in Fig. 1(a) on the constrained L_1 regression estimate with the constraint points 0.4, 0.45, \dots , 0.6 (shown in Fig. 2(a)), by deleting each observation as discussed in Section 6. For each observation number $t = 1, \dots, 50$, we compute the (unscaled) Cook distance $V_t = \|X\tilde{\beta} - X\tilde{\beta}^{\{-t\}}\|$ using both the L_2 norm and L_1 norm in the definition. The results are plotted in Fig. 7. Note that for this data set the most influential observations occur near the middle and near the right hand end point of the interval $[0, 1]$. Using Theorems 6.1 and 6.2 we find that the L_1 regression estimate remains unchanged precisely if we delete one of the seven observations, numbers 1, 2, 4, 10, 11, 16 or 17, which is consistent with Fig. 7.

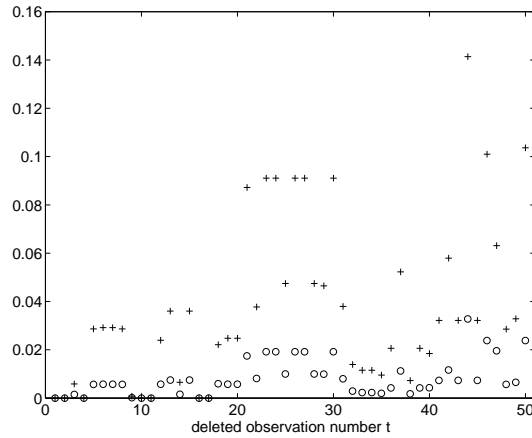


Fig. 8.7 Values of V_t defined by L_2 (o) and L_1 (+) norms

References

Anderssen, R.S. and Bloomfield, P. (1974). Numerical differentiation procedures for non-exact data, *Numer.*

- Math.* **22**, 157-182.
- Barrodale, I. and Roberts F.D.K. (1978). An efficient algorithm for discrete l_1 linear approximation with linear constraints, *SIAM J. Numer. Anal.* **15**, 603-611.
- Bloomfield, P. and Steiger, W.L. (1983). *Least Absolute Deviations, Theory, Applications and Algorithms*, Birkhäuser, Boston.
- Castillo, E., Hadi, A.S. and Lacruz, B. (2001). Regression diagnostics for the least absolute deviations and the minimax methods, *Commun. in Statist.* **30**, 1197-1225.
- Cook, R.D. (1977). Detection of influential observation in linear regression, *Technometrics* **19**, 15-18.
- Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall, New York.
- Dupačová, J. (1992). Robustness of L_1 regression in the light of linear programming, in *L_1 -Statistical Analysis and Related Methods*, ed. Dodge, Y., North-Holland, Amsterdam, 47-61.
- Effron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, New York.
- Ellis, S.P. (1998). Instability of least squares, least absolute deviation and least median of squares linear regression, *Statist. Sci.* **13**, 337-350.
- Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York.
- Gasser, T., Müller, H.G., Köhler, W., Molinari, L., and Prader, A. (1984). Nonparametric regression analysis of growth curves, *Ann. Statist.* **12**, 210-224.
- Judge, G. G., Griffiths, W. E., Carter Hill, R., Lütkepohl, H. and Tsoung-Chao Lee (1985). *The Theory and Practice of Econometrics* 2nd ed., Wiley, New York.
- Narula, S.C. and Wellington, J.F. (1985). Interior analysis for the minimum sum of absolute errors regression, *Technometrics* **27**, 181-188.
- O'Leary, D. P. and Rust, B. W. (1986). Confidence intervals for inequality-constrained least squares problems, with applications to ill-posed problems, *SIAM J. Sci. Statist. Comput.* **7**, 473-489.
- Osborne, M.R. (1985). *Finite Algorithms in Optimization and Data Analysis*, Wiley, Chichester.
- Osborne, M.R. (1987). The reduced gradient algorithm, in *Statistical Data Analysis Based on the L_1 -Norm and Related Methods*, ed. Dodge, Y., North-Holland, Amsterdam, 95-107.
- Shi, M. (1997). Constrained regularization with the L_1 norm for ill-posed problems, PhD thesis, Mathematics and Statistics, Murdoch University.
- Shi, M. and Lukas, M.A. (2002). An L_1 estimation algorithm with degeneracy and linear constraints, *Comp. Statist. Data Anal.* **39**, 35-55.
- Shi, M. and Lukas, M.A. (2005). Sensitivity analysis of constrained linear L_1 regression: perturbations to response and predictor variables, *Comp. Statist. Data Anal.* **48**, 779-802.
- Song, K.S., Müller, H.G., Clifford, A.J., Furr, H.C., and Olson, J.A. (1995). Estimating derivatives of pharmacokinetic response curves with varying bandwidths, *Biometrics* **51**, 12-20.
- Wahba, G. (1982). Constrained regularization for ill-posed linear operator equations, with applications in meteorology and medicine, in *Statistical Decision Theory and Related Topics*, III, Vol. 2, ed. Gupta, S. and Berger, J., Academic Press, New York, 383-418.
- Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia.